Lab 4

Questions with a (*) next to them form part of homework 1. Due in 20/10/2014. [Q 3]

1. What is a statistic?

2. In R, load the iris dataset, and draw boxplots for each variable
```
data(iris)
boxplot(iris)
```
What do the different lines and circles represent?

3. [*] In R, load the faithful dataset (Waiting time between eruptions and the duration of the eruption for the Old Faithful geyser in Yellowstone National Park, Wyoming, USA). Look at the eruption times `r=faithful$eruptions`.

   (a) Draw a QQ-plot. Does the data look normally distributed.

   (b) Draw a QQ-plot for only those eruptions that last longer than 3.2 minutes.

   (c) Draw a QQ-plot for only those eruptions that less than 3.2 minutes.

   (d) What do you conclude about the data?

4. In R, let $n \geq 2$ denote an integer. Draw n samples from the $N(\mu, \sigma^2)$ distribution and calculate the statistic
$$\frac{\bar{X} - \mu}{S/\sqrt{n}}$$
Repeat this N times (with perhaps $n = 5$, $N = 10^4$). Compare the empirical distribution of the N sample means with the $N(0, 1)$ distribution and the t_{n-1} distribution. How large does n have to be before the distributions are hard to distinguish?

5. Consider drawing a sample of size n from Uniform[0,1] distribution and calculate the sample median.

   (a) Does the sample median converge in probability? If so, to what?

   (b) Does the sample median converge almost surely if you increase n by adding items to the sample one by one? [Hint: If a sequence of (non-necessarily independent) events $A_1, A_2, \ldots$ have probabilities $\mathbb{P}(A_n) \leq n^{-2}$ then only finitely many of the A_n will occur (almost surely).]

6. Derive the formula for the pdf of the k-th order statistic
$$f_{X_{(j)}}(x) = \frac{n!}{(j-1)!(n-j)!} f_X(x) \left[F_X(x)\right]^{j-1} \left[1 - F_X(x)\right]^{n-j}.$$
from the cdf
$$F_{X_{(j)}}(x) = \sum_{k=j}^{n} \binom{n}{k} \left[F_X(x)\right]^k \left[1 - F_X(x)\right]^{n-k}$$

7. Consider the Poisson(θ) distribution.

   (a) In R, sample 4 items with $\theta = 2.7$.

   (b) Plot the likelihood for θ in the range one to five.

   (c) Can you explain the position of the maximum in terms of the sample mean?

   (d) In R, sample 100 items with $\theta = 2.7$.

   (e) Plot the log of the likelihood for θ in the range one to five.

   (f) Can you explain the position of the maximum in terms of the sample mean?

Lab 4

Questions with a (*) next to them form part of homework 1. Due in 20/10/2014. [Q 3]

1. What is a statistic?

2. In R, load the iris dataset, and draw boxplots for each variable
```
data(iris)
boxplot(iris)
```
   What do the different lines and circles represent?

3. [*] In R, load the faithful dataset (Waiting time between eruptions and the duration of the eruption for the Old Faithful geyser in Yellowstone National Park, Wyoming, USA). Look at the eruption times `r=faithful$eruptions`.

   (a) Draw a QQ-plot. Does the data look normally distributed.

   (b) Draw a QQ-plot for only those eruptions that last longer than 3.2 minutes.

   (c) Draw a QQ-plot for only those eruptions that less than 3.2 minutes.

   (d) What do you conclude about the data?

4. In R, let $n \geq 2$ denote an integer. Draw n samples from the $N(\mu, \sigma^2)$ distribution and calculate the statistic

$$\frac{\bar{X} - \mu}{S/\sqrt{n}}$$

   Repeat this N times (with perhaps $n = 5$, $N = 10^4$). Compare the empirical distribution of the N sample means with the $N(0, 1)$ distribution and the t_{n-1} distribution. How large does n have to be before the distributions are hard to distinguish?

5. Consider drawing a sample of size n from Uniform[0,1] distribution and calculate the sample median.

   (a) Does the sample median converge in probability? If so, to what?

   (b) Does the sample median converge almost surely if you increase n by adding items to the sample one by one? [Hint: If a sequence of (non-necessarily independent) events $A_1, A_2, \ldots$ have probabilities $\mathbb{P}(A_n) \leq n^{-2}$ then only finitely many of the A_n will occur (almost surely).]

6. Derive the formula for the pdf of the k-th order statistic

$$f_{X_{(j)}}(x) = \frac{n!}{(j-1)!(n-j)!} f_X(x) \left[F_X(x)\right]^{j-1} \left[1 - F_X(x)\right]^{n-j}.$$

   from the cdf

$$F_{X_{(j)}}(x) = \sum_{k=j}^{n} \binom{n}{k} \left[F_X(x)\right]^k \left[1 - F_X(x)\right]^{n-k}$$

7. Consider the Poisson(θ) distribution.

   (a) In R, sample 4 items with $\theta = 2.7$.

   (b) Plot the likelihood for θ in the range one to five.

   (c) Can you explain the position of the maximum in terms of the sample mean?

   (d) In R, sample 100 items with $\theta = 2.7$.

   (e) Plot the log of the likelihood for θ in the range one to five.

   (f) Can you explain the position of the maximum in terms of the sample mean?

Lab 4

Questions with a (*) next to them form part of homework 1. Due in 20/10/2014. [Q 3]

1. What is a statistic?

2. In R, load the iris dataset, and draw boxplots for each variable
```
data(iris)
boxplot(iris)
```
   What do the different lines and circles represent?

3. [*] In R, load the faithful dataset (Waiting time between eruptions and the duration of the eruption for the Old Faithful geyser in Yellowstone National Park, Wyoming, USA). Look at the eruption times `r=faithful$eruptions`.

   (a) Draw a QQ-plot. Does the data look normally distributed.

   (b) Draw a QQ-plot for only those eruptions that last longer than 3.2 minutes.

   (c) Draw a QQ-plot for only those eruptions that less than 3.2 minutes.

   (d) What do you conclude about the data?

4. In R, let $n \geq 2$ denote an integer. Draw n samples from the $N(\mu, \sigma^2)$ distribution and calculate the statistic
$$\frac{\bar{X} - \mu}{S/\sqrt{n}}$$
   Repeat this N times (with perhaps $n = 5$, $N = 10^4$). Compare the empirical distribution of the N sample means with the $N(0, 1)$ distribution and the t_{n-1} distribution. How large does n have to be before the distributions are hard to distinguish?

5. Consider drawing a sample of size n from Uniform[0,1] distribution and calculate the sample median.

   (a) Does the sample median converge in probability? If so, to what?

   (b) Does the sample median converge almost surely if you increase n by adding items to the sample one by one? [Hint: If a sequence of (non-necessarily independent) events $A_1, A_2, \ldots$ have probabilities $\mathbb{P}(A_n) \leq n^{-2}$ then only finitely many of the A_n will occur (almost surely).]

6. Derive the formula for the pdf of the k-th order statistic
$$f_{X_{(j)}}(x) = \frac{n!}{(j-1)!(n-j)!} f_X(x) \left[F_X(x)\right]^{j-1} \left[1 - F_X(x)\right]^{n-j}.$$
   from the cdf
$$F_{X_{(j)}}(x) = \sum_{k=j}^{n} \binom{n}{k} \left[F_X(x)\right]^k \left[1 - F_X(x)\right]^{n-k}$$

7. Consider the Poisson(θ) distribution.

   (a) In R, sample 4 items with $\theta = 2.7$.

   (b) Plot the likelihood for θ in the range one to five.

   (c) Can you explain the position of the maximum in terms of the sample mean?

   (d) In R, sample 100 items with $\theta = 2.7$.

   (e) Plot the log of the likelihood for θ in the range one to five.

   (f) Can you explain the position of the maximum in terms of the sample mean?

Lab 4

Questions with a (*) next to them form part of homework 1. Due in 20/10/2014. [Q 3]

1. What is a statistic?

2. In R, load the iris dataset, and draw boxplots for each variable
```
data(iris)
boxplot(iris)
```
What do the different lines and circles represent?

3. [*] In R, load the faithful dataset (Waiting time between eruptions and the duration of the eruption for the Old Faithful geyser in Yellowstone National Park, Wyoming, USA). Look at the eruption times `r=faithful$eruptions`.

   (a) Draw a QQ-plot. Does the data look normally distributed.

   (b) Draw a QQ-plot for only those eruptions that last longer than 3.2 minutes.

   (c) Draw a QQ-plot for only those eruptions that less than 3.2 minutes.

   (d) What do you conclude about the data?

4. In R, let $n \geq 2$ denote an integer. Draw n samples from the $N(\mu, \sigma^2)$ distribution and calculate the statistic
$$\frac{\bar{X} - \mu}{S/\sqrt{n}}$$
Repeat this N times (with perhaps $n = 5$, $N = 10^4$). Compare the empirical distribution of the N sample means with the $N(0, 1)$ distribution and the t_{n-1} distribution. How large does n have to be before the distributions are hard to distinguish?

5. Consider drawing a sample of size n from Uniform[0,1] distribution and calculate the sample median.

   (a) Does the sample median converge in probability? If so, to what?

   (b) Does the sample median converge almost surely if you increase n by adding items to the sample one by one? [Hint: If a sequence of (non-necessarily independent) events $A_1, A_2, \ldots$ have probabilities $\mathbb{P}(A_n) \leq n^{-2}$ then only finitely many of the A_n will occur (almost surely).]

6. Derive the formula for the pdf of the k-th order statistic
$$f_{X_{(j)}}(x) = \frac{n!}{(j-1)!(n-j)!} f_X(x) \left[F_X(x)\right]^{j-1} \left[1 - F_X(x)\right]^{n-j}.$$
from the cdf
$$F_{X_{(j)}}(x) = \sum_{k=j}^{n} \binom{n}{k} \left[F_X(x)\right]^{k} \left[1 - F_X(x)\right]^{n-k}$$

7. Consider the Poisson(θ) distribution.

   (a) In R, sample 4 items with $\theta = 2.7$.

   (b) Plot the likelihood for θ in the range one to five.

   (c) Can you explain the position of the maximum in terms of the sample mean?

   (d) In R, sample 100 items with $\theta = 2.7$.

   (e) Plot the log of the likelihood for θ in the range one to five.

   (f) Can you explain the position of the maximum in terms of the sample mean?

Lab 4

Questions with a (*) next to them form part of homework 1. Due in 20/10/2014. [Q 3]

1. What is a statistic?

2. In R, load the iris dataset, and draw boxplots for each variable
```
data(iris)
boxplot(iris)
```
What do the different lines and circles represent?

3. [*] In R, load the faithful dataset (Waiting time between eruptions and the duration of the eruption for the Old Faithful geyser in Yellowstone National Park, Wyoming, USA). Look at the eruption times `r=faithful$eruptions`.

   (a) Draw a QQ-plot. Does the data look normally distributed.

   (b) Draw a QQ-plot for only those eruptions that last longer than 3.2 minutes.

   (c) Draw a QQ-plot for only those eruptions that less than 3.2 minutes.

   (d) What do you conclude about the data?

4. In R, let $n \geq 2$ denote an integer. Draw n samples from the $N(\mu, \sigma^2)$ distribution and calculate the statistic
$$\frac{\bar{X} - \mu}{S/\sqrt{n}}$$
Repeat this N times (with perhaps $n = 5$, $N = 10^4$). Compare the empirical distribution of the N sample means with the $N(0, 1)$ distribution and the t_{n-1} distribution. How large does n have to be before the distributions are hard to distinguish?

5. Consider drawing a sample of size n from Uniform[0,1] distribution and calculate the sample median.

   (a) Does the sample median converge in probability? If so, to what?

   (b) Does the sample median converge almost surely if you increase n by adding items to the sample one by one? [Hint: If a sequence of (non-necessarily independent) events $A_1, A_2, \ldots$ have probabilities $\mathbb{P}(A_n) \leq n^{-2}$ then only finitely many of the A_n will occur (almost surely).]

6. Derive the formula for the pdf of the k-th order statistic
$$f_{X_{(j)}}(x) = \frac{n!}{(j-1)!(n-j)!} f_X(x) \left[F_X(x)\right]^{j-1} \left[1 - F_X(x)\right]^{n-j}.$$
from the cdf
$$F_{X_{(j)}}(x) = \sum_{k=j}^{n} \binom{n}{k} \left[F_X(x)\right]^{k} \left[1 - F_X(x)\right]^{n-k}$$

7. Consider the Poisson(θ) distribution.

   (a) In R, sample 4 items with $\theta = 2.7$.

   (b) Plot the likelihood for θ in the range one to five.

   (c) Can you explain the position of the maximum in terms of the sample mean?

   (d) In R, sample 100 items with $\theta = 2.7$.

   (e) Plot the log of the likelihood for θ in the range one to five.

   (f) Can you explain the position of the maximum in terms of the sample mean?

Lab 4

Questions with a (*) next to them form part of homework 1. Due in 20/10/2014. [Q 3]

1. What is a statistic?

2. In R, load the iris dataset, and draw boxplots for each variable
   `data(iris)`
   `boxplot(iris)`
   What do the different lines and circles represent?

3. [*] In R, load the faithful dataset (Waiting time between eruptions and the duration of the eruption for the Old Faithful geyser in Yellowstone National Park, Wyoming, USA). Look at the eruption times `r=faithful$eruptions`.

   (a) Draw a QQ-plot. Does the data look normally distributed.

   (b) Draw a QQ-plot for only those eruptions that last longer than 3.2 minutes.

   (c) Draw a QQ-plot for only those eruptions that less than 3.2 minutes.

   (d) What do you conclude about the data?

4. In R, let $n \geq 2$ denote an integer. Draw n samples from the $N(\mu, \sigma^2)$ distribution and calculate the statistic

$$\frac{\bar{X} - \mu}{S/\sqrt{n}}$$

   Repeat this N times (with perhaps $n = 5$, $N = 10^4$). Compare the empirical distribution of the N sample means with the $N(0, 1)$ distribution and the t_{n-1} distribution. How large does n have to be before the distributions are hard to distinguish?

5. Consider drawing a sample of size n from Uniform[0,1] distribution and calculate the sample median.

   (a) Does the sample median converge in probability? If so, to what?

   (b) Does the sample median converge almost surely if you increase n by adding items to the sample one by one? [Hint: If a sequence of (non-necessarily independent) events $A_1, A_2, \ldots$ have probabilities $\mathbb{P}(A_n) \leq n^{-2}$ then only finitely many of the A_n will occur (almost surely).]

6. Derive the formula for the pdf of the k-th order statistic

$$f_{X_{(j)}}(x) = \frac{n!}{(j-1)!(n-j)!} f_X(x) \left[F_X(x)\right]^{j-1} \left[1 - F_X(x)\right]^{n-j}.$$

   from the cdf

$$F_{X_{(j)}}(x) = \sum_{k=j}^{n} \binom{n}{k} \left[F_X(x)\right]^k \left[1 - F_X(x)\right]^{n-k}$$

7. Consider the Poisson(θ) distribution.

   (a) In R, sample 4 items with $\theta = 2.7$.

   (b) Plot the likelihood for θ in the range one to five.

   (c) Can you explain the position of the maximum in terms of the sample mean?

   (d) In R, sample 100 items with $\theta = 2.7$.

   (e) Plot the log of the likelihood for θ in the range one to five.

   (f) Can you explain the position of the maximum in terms of the sample mean?

Lab 4

Questions with a (*) next to them form part of homework 1. Due in 20/10/2014. [Q 3]

1. What is a statistic?
2. In R, load the iris dataset, and draw boxplots for each variable
```
data(iris)
boxplot(iris)
```
What do the different lines and circles represent?
3. [*] In R, load the faithful dataset (Waiting time between eruptions and the duration of the eruption for the Old Faithful geyser in Yellowstone National Park, Wyoming, USA). Look at the eruption times `r=faithful$eruptions`.
   (a) Draw a QQ-plot. Does the data look normally distributed.
   (b) Draw a QQ-plot for only those eruptions that last longer than 3.2 minutes.
   (c) Draw a QQ-plot for only those eruptions that less than 3.2 minutes.
   (d) What do you conclude about the data?
4. In R, let $n \geq 2$ denote an integer. Draw n samples from the $N(\mu, \sigma^2)$ distribution and calculate the statistic
$$\frac{\bar{X} - \mu}{S/\sqrt{n}}$$
Repeat this N times (with perhaps $n = 5$, $N = 10^4$). Compare the empirical distribution of the N sample means with the $N(0, 1)$ distribution and the t_{n-1} distribution. How large does n have to be before the distributions are hard to distinguish?
5. Consider drawing a sample of size n from Uniform[0,1] distribution and calculate the sample median.
   (a) Does the sample median converge in probability? If so, to what?
   (b) Does the sample median converge almost surely if you increase n by adding items to the sample one by one? [Hint: If a sequence of (non-necessarily independent) events $A_1, A_2, \ldots$ have probabilities $\mathbb{P}(A_n) \leq n^{-2}$ then only finitely many of the A_n will occur (almost surely).]
6. Derive the formula for the pdf of the k-th order statistic
$$f_{X_{(j)}}(x) = \frac{n!}{(j-1)!(n-j)!} f_X(x) \left[F_X(x)\right]^{j-1} \left[1 - F_X(x)\right]^{n-j}.$$
from the cdf
$$F_{X_{(j)}}(x) = \sum_{k=j}^{n} \binom{n}{k} \left[F_X(x)\right]^k \left[1 - F_X(x)\right]^{n-k}$$
7. Consider the Poisson(θ) distribution.
   (a) In R, sample 4 items with $\theta = 2.7$.
   (b) Plot the likelihood for θ in the range one to five.
   (c) Can you explain the position of the maximum in terms of the sample mean?
   (d) In R, sample 100 items with $\theta = 2.7$.
   (e) Plot the log of the likelihood for θ in the range one to five.
   (f) Can you explain the position of the maximum in terms of the sample mean?

Lab 4

Questions with a (*) next to them form part of homework 1. Due in 20/10/2014. [Q 3]

1. What is a statistic?

2. In R, load the iris dataset, and draw boxplots for each variable
```
data(iris)
boxplot(iris)
```
What do the different lines and circles represent?

3. [*] In R, load the faithful dataset (Waiting time between eruptions and the duration of the eruption for the Old Faithful geyser in Yellowstone National Park, Wyoming, USA). Look at the eruption times `r=faithful$eruptions`.

   (a) Draw a QQ-plot. Does the data look normally distributed.

   (b) Draw a QQ-plot for only those eruptions that last longer than 3.2 minutes.

   (c) Draw a QQ-plot for only those eruptions that less than 3.2 minutes.

   (d) What do you conclude about the data?

4. In R, let $n \geq 2$ denote an integer. Draw n samples from the $N(\mu, \sigma^2)$ distribution and calculate the statistic
$$\frac{\bar{X} - \mu}{S/\sqrt{n}}$$
Repeat this N times (with perhaps $n = 5$, $N = 10^4$). Compare the empirical distribution of the N sample means with the $N(0,1)$ distribution and the t_{n-1} distribution. How large does n have to be before the distributions are hard to distinguish?

5. Consider drawing a sample of size n from Uniform[0,1] distribution and calculate the sample median.

   (a) Does the sample median converge in probability? If so, to what?

   (b) Does the sample median converge almost surely if you increase n by adding items to the sample one by one? [Hint: If a sequence of (non-necessarily independent) events $A_1, A_2, \ldots$ have probabilities $\mathbb{P}(A_n) \leq n^{-2}$ then only finitely many of the A_n will occur (almost surely).]

6. Derive the formula for the pdf of the k-th order statistic
$$f_{X_{(j)}}(x) = \frac{n!}{(j-1)!(n-j)!} f_X(x) \left[F_X(x)\right]^{j-1} \left[1 - F_X(x)\right]^{n-j}.$$
from the cdf
$$F_{X_{(j)}}(x) = \sum_{k=j}^{n} \binom{n}{k} \left[F_X(x)\right]^k \left[1 - F_X(x)\right]^{n-k}$$

7. Consider the Poisson(θ) distribution.

   (a) In R, sample 4 items with $\theta = 2.7$.

   (b) Plot the likelihood for θ in the range one to five.

   (c) Can you explain the position of the maximum in terms of the sample mean?

   (d) In R, sample 100 items with $\theta = 2.7$.

   (e) Plot the log of the likelihood for θ in the range one to five.

   (f) Can you explain the position of the maximum in terms of the sample mean?

Lab 4

Questions with a (*) next to them form part of homework 1. Due in 20/10/2014. [Q 3]

1. What is a statistic?

2. In R, load the iris dataset, and draw boxplots for each variable
```
data(iris)
boxplot(iris)
```
What do the different lines and circles represent?

3. [*] In R, load the faithful dataset (Waiting time between eruptions and the duration of the eruption for the Old Faithful geyser in Yellowstone National Park, Wyoming, USA). Look at the eruption times `r=faithful$eruptions`.

   (a) Draw a QQ-plot. Does the data look normally distributed.

   (b) Draw a QQ-plot for only those eruptions that last longer than 3.2 minutes.

   (c) Draw a QQ-plot for only those eruptions that less than 3.2 minutes.

   (d) What do you conclude about the data?

4. In R, let $n \geq 2$ denote an integer. Draw n samples from the $N(\mu, \sigma^2)$ distribution and calculate the statistic
$$\frac{\bar{X} - \mu}{S/\sqrt{n}}$$
Repeat this N times (with perhaps $n = 5$, $N = 10^4$). Compare the empirical distribution of the N sample means with the $N(0, 1)$ distribution and the t_{n-1} distribution. How large does n have to be before the distributions are hard to distinguish?

5. Consider drawing a sample of size n from Uniform[0,1] distribution and calculate the sample median.

   (a) Does the sample median converge in probability? If so, to what?

   (b) Does the sample median converge almost surely if you increase n by adding items to the sample one by one? [Hint: If a sequence of (non-necessarily independent) events $A_1, A_2, \dots$ have probabilities $\mathbb{P}(A_n) \leq n^{-2}$ then only finitely many of the A_n will occur (almost surely).]

6. Derive the formula for the pdf of the k-th order statistic
$$f_{X_{(j)}}(x) = \frac{n!}{(j-1)!(n-j)!} f_X(x) \left[F_X(x)\right]^{j-1} \left[1 - F_X(x)\right]^{n-j}.$$
from the cdf
$$F_{X_{(j)}}(x) = \sum_{k=j}^{n} \binom{n}{k} \left[F_X(x)\right]^k \left[1 - F_X(x)\right]^{n-k}$$

7. Consider the Poisson(θ) distribution.

   (a) In R, sample 4 items with $\theta = 2.7$.

   (b) Plot the likelihood for θ in the range one to five.

   (c) Can you explain the position of the maximum in terms of the sample mean?

   (d) In R, sample 100 items with $\theta = 2.7$.

   (e) Plot the log of the likelihood for θ in the range one to five.

   (f) Can you explain the position of the maximum in terms of the sample mean?

Lab 4

Questions with a (*) next to them form part of homework 1. Due in 20/10/2014. [Q 3]

1. What is a statistic?

2. In R, load the iris dataset, and draw boxplots for each variable
```
data(iris)
boxplot(iris)
```
What do the different lines and circles represent?

3. [*] In R, load the faithful dataset (Waiting time between eruptions and the duration of the eruption for the Old Faithful geyser in Yellowstone National Park, Wyoming, USA). Look at the eruption times `r=faithful$eruptions`.

   (a) Draw a QQ-plot. Does the data look normally distributed.

   (b) Draw a QQ-plot for only those eruptions that last longer than 3.2 minutes.

   (c) Draw a QQ-plot for only those eruptions that less than 3.2 minutes.

   (d) What do you conclude about the data?

4. In R, let $n \geq 2$ denote an integer. Draw n samples from the $N(\mu, \sigma^2)$ distribution and calculate the statistic
$$\frac{\bar{X} - \mu}{S/\sqrt{n}}$$
Repeat this N times (with perhaps $n = 5$, $N = 10^4$). Compare the empirical distribution of the N sample means with the $N(0, 1)$ distribution and the t_{n-1} distribution. How large does n have to be before the distributions are hard to distinguish?

5. Consider drawing a sample of size n from Uniform[0,1] distribution and calculate the sample median.

   (a) Does the sample median converge in probability? If so, to what?

   (b) Does the sample median converge almost surely if you increase n by adding items to the sample one by one? [Hint: If a sequence of (non-necessarily independent) events $A_1, A_2, \ldots$ have probabilities $\mathbb{P}(A_n) \leq n^{-2}$ then only finitely many of the A_n will occur (almost surely).]

6. Derive the formula for the pdf of the k-th order statistic
$$f_{X_{(j)}}(x) = \frac{n!}{(j-1)!(n-j)!} f_X(x) \left[F_X(x)\right]^{j-1} \left[1 - F_X(x)\right]^{n-j}.$$
from the cdf
$$F_{X_{(j)}}(x) = \sum_{k=j}^{n} \binom{n}{k} \left[F_X(x)\right]^k \left[1 - F_X(x)\right]^{n-k}$$

7. Consider the Poisson(θ) distribution.

   (a) In R, sample 4 items with $\theta = 2.7$.

   (b) Plot the likelihood for θ in the range one to five.

   (c) Can you explain the position of the maximum in terms of the sample mean?

   (d) In R, sample 100 items with $\theta = 2.7$.

   (e) Plot the log of the likelihood for θ in the range one to five.

   (f) Can you explain the position of the maximum in terms of the sample mean?

Lab 4

Questions with a (*) next to them form part of homework 1. Due in 20/10/2014. [Q 3]

1. What is a statistic?

2. In R, load the iris dataset, and draw boxplots for each variable
```
data(iris)
boxplot(iris)
```
What do the different lines and circles represent?

3. [*] In R, load the faithful dataset (Waiting time between eruptions and the duration of the eruption for the Old Faithful geyser in Yellowstone National Park, Wyoming, USA). Look at the eruption times `r=faithful$eruptions`.

   (a) Draw a QQ-plot. Does the data look normally distributed.

   (b) Draw a QQ-plot for only those eruptions that last longer than 3.2 minutes.

   (c) Draw a QQ-plot for only those eruptions that less than 3.2 minutes.

   (d) What do you conclude about the data?

4. In R, let $n \geq 2$ denote an integer. Draw n samples from the $N(\mu, \sigma^2)$ distribution and calculate the statistic
$$\frac{\bar{X} - \mu}{S/\sqrt{n}}$$
Repeat this N times (with perhaps $n = 5$, $N = 10^4$). Compare the empirical distribution of the N sample means with the $N(0,1)$ distribution and the t_{n-1} distribution. How large does n have to be before the distributions are hard to distinguish?

5. Consider drawing a sample of size n from Uniform[0,1] distribution and calculate the sample median.

   (a) Does the sample median converge in probability? If so, to what?

   (b) Does the sample median converge almost surely if you increase n by adding items to the sample one by one? [Hint: If a sequence of (non-necessarily independent) events $A_1, A_2, \ldots$ have probabilities $\mathbb{P}(A_n) \leq n^{-2}$ then only finitely many of the A_n will occur (almost surely).]

6. Derive the formula for the pdf of the k-th order statistic
$$f_{X_{(j)}}(x) = \frac{n!}{(j-1)!(n-j)!} f_X(x) \left[F_X(x)\right]^{j-1} \left[1 - F_X(x)\right]^{n-j}.$$
from the cdf
$$F_{X_{(j)}}(x) = \sum_{k=j}^{n} \binom{n}{k} \left[F_X(x)\right]^k \left[1 - F_X(x)\right]^{n-k}$$

7. Consider the Poisson(θ) distribution.

   (a) In R, sample 4 items with $\theta = 2.7$.

   (b) Plot the likelihood for θ in the range one to five.

   (c) Can you explain the position of the maximum in terms of the sample mean?

   (d) In R, sample 100 items with $\theta = 2.7$.

   (e) Plot the log of the likelihood for θ in the range one to five.

   (f) Can you explain the position of the maximum in terms of the sample mean?

Lab 4

Questions with a (*) next to them form part of homework 1. Due in 20/10/2014. [Q 3]

1. What is a statistic?

2. In R, load the iris dataset, and draw boxplots for each variable
```
data(iris)
boxplot(iris)
```
   What do the different lines and circles represent?

3. [*] In R, load the faithful dataset (Waiting time between eruptions and the duration of the eruption for the Old Faithful geyser in Yellowstone National Park, Wyoming, USA). Look at the eruption times `r=faithful$eruptions`.

   (a) Draw a QQ-plot. Does the data look normally distributed.

   (b) Draw a QQ-plot for only those eruptions that last longer than 3.2 minutes.

   (c) Draw a QQ-plot for only those eruptions that less than 3.2 minutes.

   (d) What do you conclude about the data?

4. In R, let $n \geq 2$ denote an integer. Draw n samples from the $N(\mu, \sigma^2)$ distribution and calculate the statistic
$$\frac{\bar{X} - \mu}{S/\sqrt{n}}$$
   Repeat this N times (with perhaps $n = 5$, $N = 10^4$). Compare the empirical distribution of the N sample means with the $N(0, 1)$ distribution and the t_{n-1} distribution. How large does n have to be before the distributions are hard to distinguish?

5. Consider drawing a sample of size n from Uniform[0,1] distribution and calculate the sample median.

   (a) Does the sample median converge in probability? If so, to what?

   (b) Does the sample median converge almost surely if you increase n by adding items to the sample one by one? [Hint: If a sequence of (non-necessarily independent) events $A_1, A_2, \ldots$ have probabilities $\mathbb{P}(A_n) \leq n^{-2}$ then only finitely many of the A_n will occur (almost surely).]

6. Derive the formula for the pdf of the k-th order statistic
$$f_{X_{(j)}}(x) = \frac{n!}{(j-1)!(n-j)!} f_X(x) \left[F_X(x)\right]^{j-1} \left[1 - F_X(x)\right]^{n-j}.$$
   from the cdf
$$F_{X_{(j)}}(x) = \sum_{k=j}^{n} \binom{n}{k} \left[F_X(x)\right]^k \left[1 - F_X(x)\right]^{n-k}$$

7. Consider the Poisson(θ) distribution.

   (a) In R, sample 4 items with $\theta = 2.7$.

   (b) Plot the likelihood for θ in the range one to five.

   (c) Can you explain the position of the maximum in terms of the sample mean?

   (d) In R, sample 100 items with $\theta = 2.7$.

   (e) Plot the log of the likelihood for θ in the range one to five.

   (f) Can you explain the position of the maximum in terms of the sample mean?

Lab 4

Questions with a (*) next to them form part of homework 1. Due in 20/10/2014. [Q 3]

1. What is a statistic?

2. In R, load the iris dataset, and draw boxplots for each variable
```
data(iris)
boxplot(iris)
```
What do the different lines and circles represent?

3. [*] In R, load the faithful dataset (Waiting time between eruptions and the duration of the eruption for the Old Faithful geyser in Yellowstone National Park, Wyoming, USA). Look at the eruption times `r=faithful$eruptions`.

   (a) Draw a QQ-plot. Does the data look normally distributed.

   (b) Draw a QQ-plot for only those eruptions that last longer than 3.2 minutes.

   (c) Draw a QQ-plot for only those eruptions that less than 3.2 minutes.

   (d) What do you conclude about the data?

4. In R, let $n \geq 2$ denote an integer. Draw n samples from the $N(\mu, \sigma^2)$ distribution and calculate the statistic
$$\frac{\bar{X} - \mu}{S/\sqrt{n}}$$
Repeat this N times (with perhaps $n = 5$, $N = 10^4$). Compare the empirical distribution of the N sample means with the $N(0, 1)$ distribution and the t_{n-1} distribution. How large does n have to be before the distributions are hard to distinguish?

5. Consider drawing a sample of size n from Uniform[0,1] distribution and calculate the sample median.

   (a) Does the sample median converge in probability? If so, to what?

   (b) Does the sample median converge almost surely if you increase n by adding items to the sample one by one? [Hint: If a sequence of (non-necessarily independent) events $A_1, A_2, \dots$ have probabilities $\mathbb{P}(A_n) \leq n^{-2}$ then only finitely many of the A_n will occur (almost surely).]

6. Derive the formula for the pdf of the k-th order statistic
$$f_{X_{(j)}}(x) = \frac{n!}{(j-1)!(n-j)!} f_X(x) \left[F_X(x)\right]^{j-1} \left[1 - F_X(x)\right]^{n-j}.$$
from the cdf
$$F_{X_{(j)}}(x) = \sum_{k=j}^{n} \binom{n}{k} \left[F_X(x)\right]^{k} \left[1 - F_X(x)\right]^{n-k}$$

7. Consider the Poisson(θ) distribution.

   (a) In R, sample 4 items with $\theta = 2.7$.

   (b) Plot the likelihood for θ in the range one to five.

   (c) Can you explain the position of the maximum in terms of the sample mean?

   (d) In R, sample 100 items with $\theta = 2.7$.

   (e) Plot the log of the likelihood for θ in the range one to five.

   (f) Can you explain the position of the maximum in terms of the sample mean?

Lab 4

Questions with a (*) next to them form part of homework 1. Due in 20/10/2014. [Q 3]

1. What is a statistic?

2. In R, load the iris dataset, and draw boxplots for each variable
```
data(iris)
boxplot(iris)
```
What do the different lines and circles represent?

3. [*] In R, load the faithful dataset (Waiting time between eruptions and the duration of the eruption for the Old Faithful geyser in Yellowstone National Park, Wyoming, USA). Look at the eruption times `r=faithful$eruptions`.

   (a) Draw a QQ-plot. Does the data look normally distributed.

   (b) Draw a QQ-plot for only those eruptions that last longer than 3.2 minutes.

   (c) Draw a QQ-plot for only those eruptions that less than 3.2 minutes.

   (d) What do you conclude about the data?

4. In R, let $n \geq 2$ denote an integer. Draw n samples from the $N(\mu, \sigma^2)$ distribution and calculate the statistic
$$\frac{\bar{X} - \mu}{S/\sqrt{n}}$$
Repeat this N times (with perhaps $n = 5$, $N = 10^4$). Compare the empirical distribution of the N sample means with the $N(0,1)$ distribution and the t_{n-1} distribution. How large does n have to be before the distributions are hard to distinguish?

5. Consider drawing a sample of size n from Uniform[0,1] distribution and calculate the sample median.

   (a) Does the sample median converge in probability? If so, to what?

   (b) Does the sample median converge almost surely if you increase n by adding items to the sample one by one? [Hint: If a sequence of (non-necessarily independent) events $A_1, A_2, \ldots$ have probabilities $\mathbb{P}(A_n) \leq n^{-2}$ then only finitely many of the A_n will occur (almost surely).]

6. Derive the formula for the pdf of the k-th order statistic
$$f_{X_{(j)}}(x) = \frac{n!}{(j-1)!(n-j)!} f_X(x) \left[F_X(x)\right]^{j-1} \left[1 - F_X(x)\right]^{n-j}.$$
from the cdf
$$F_{X_{(j)}}(x) = \sum_{k=j}^{n} \binom{n}{k} \left[F_X(x)\right]^k \left[1 - F_X(x)\right]^{n-k}$$

7. Consider the Poisson(θ) distribution.

   (a) In R, sample 4 items with $\theta = 2.7$.

   (b) Plot the likelihood for θ in the range one to five.

   (c) Can you explain the position of the maximum in terms of the sample mean?

   (d) In R, sample 100 items with $\theta = 2.7$.

   (e) Plot the log of the likelihood for θ in the range one to five.

   (f) Can you explain the position of the maximum in terms of the sample mean?

Lab 4

Questions with a (*) next to them form part of homework 1. Due in 20/10/2014. [Q 3]

1. What is a statistic?

2. In R, load the iris dataset, and draw boxplots for each variable
```
data(iris)
boxplot(iris)
```
What do the different lines and circles represent?

3. [*] In R, load the faithful dataset (Waiting time between eruptions and the duration of the eruption for the Old Faithful geyser in Yellowstone National Park, Wyoming, USA). Look at the eruption times `r=faithful$eruptions`.

   (a) Draw a QQ-plot. Does the data look normally distributed.

   (b) Draw a QQ-plot for only those eruptions that last longer than 3.2 minutes.

   (c) Draw a QQ-plot for only those eruptions that less than 3.2 minutes.

   (d) What do you conclude about the data?

4. In R, let $n \geq 2$ denote an integer. Draw n samples from the $N(\mu, \sigma^2)$ distribution and calculate the statistic
$$\frac{\bar{X} - \mu}{S/\sqrt{n}}$$
Repeat this N times (with perhaps $n = 5$, $N = 10^4$). Compare the empirical distribution of the N sample means with the $N(0, 1)$ distribution and the t_{n-1} distribution. How large does n have to be before the distributions are hard to distinguish?

5. Consider drawing a sample of size n from Uniform[0,1] distribution and calculate the sample median.

   (a) Does the sample median converge in probability? If so, to what?

   (b) Does the sample median converge almost surely if you increase n by adding items to the sample one by one? [Hint: If a sequence of (non-necessarily independent) events $A_1, A_2, \ldots$ have probabilities $\mathbb{P}(A_n) \leq n^{-2}$ then only finitely many of the A_n will occur (almost surely).]

6. Derive the formula for the pdf of the k-th order statistic
$$f_{X_{(j)}}(x) = \frac{n!}{(j-1)!(n-j)!} f_X(x) \left[F_X(x)\right]^{j-1} \left[1 - F_X(x)\right]^{n-j}.$$
from the cdf
$$F_{X_{(j)}}(x) = \sum_{k=j}^{n} \binom{n}{k} \left[F_X(x)\right]^k \left[1 - F_X(x)\right]^{n-k}$$

7. Consider the Poisson(θ) distribution.

   (a) In R, sample 4 items with $\theta = 2.7$.

   (b) Plot the likelihood for θ in the range one to five.

   (c) Can you explain the position of the maximum in terms of the sample mean?

   (d) In R, sample 100 items with $\theta = 2.7$.

   (e) Plot the log of the likelihood for θ in the range one to five.

   (f) Can you explain the position of the maximum in terms of the sample mean?

Lab 4

Questions with a (*) next to them form part of homework 1. Due in 20/10/2014. [Q 3]

1. What is a statistic?

2. In R, load the iris dataset, and draw boxplots for each variable
```
data(iris)
boxplot(iris)
```
What do the different lines and circles represent?

3. [*] In R, load the faithful dataset (Waiting time between eruptions and the duration of the eruption for the Old Faithful geyser in Yellowstone National Park, Wyoming, USA). Look at the eruption times `r=faithful$eruptions`.

   (a) Draw a QQ-plot. Does the data look normally distributed.

   (b) Draw a QQ-plot for only those eruptions that last longer than 3.2 minutes.

   (c) Draw a QQ-plot for only those eruptions that less than 3.2 minutes.

   (d) What do you conclude about the data?

4. In R, let $n \geq 2$ denote an integer. Draw n samples from the $N(\mu, \sigma^2)$ distribution and calculate the statistic
$$\frac{\bar{X} - \mu}{S/\sqrt{n}}$$
Repeat this N times (with perhaps $n = 5$, $N = 10^4$). Compare the empirical distribution of the N sample means with the $N(0, 1)$ distribution and the t_{n-1} distribution. How large does n have to be before the distributions are hard to distinguish?

5. Consider drawing a sample of size n from Uniform[0,1] distribution and calculate the sample median.

   (a) Does the sample median converge in probability? If so, to what?

   (b) Does the sample median converge almost surely if you increase n by adding items to the sample one by one? [Hint: If a sequence of (non-necessarily independent) events $A_1, A_2, \ldots$ have probabilities $\mathbb{P}(A_n) \leq n^{-2}$ then only finitely many of the A_n will occur (almost surely).]

6. Derive the formula for the pdf of the k-th order statistic
$$f_{X_{(j)}}(x) = \frac{n!}{(j-1)!(n-j)!} f_X(x) \left[F_X(x)\right]^{j-1} \left[1 - F_X(x)\right]^{n-j}.$$
from the cdf
$$F_{X_{(j)}}(x) = \sum_{k=j}^{n} \binom{n}{k} \left[F_X(x)\right]^k \left[1 - F_X(x)\right]^{n-k}$$

7. Consider the Poisson(θ) distribution.

   (a) In R, sample 4 items with $\theta = 2.7$.

   (b) Plot the likelihood for θ in the range one to five.

   (c) Can you explain the position of the maximum in terms of the sample mean?

   (d) In R, sample 100 items with $\theta = 2.7$.

   (e) Plot the log of the likelihood for θ in the range one to five.

   (f) Can you explain the position of the maximum in terms of the sample mean?

Lab 4

Questions with a (*) next to them form part of homework 1. Due in 20/10/2014. [Q 3]

1. What is a statistic?

2. In R, load the iris dataset, and draw boxplots for each variable
```
data(iris)
boxplot(iris)
```
What do the different lines and circles represent?

3. [*] In R, load the faithful dataset (Waiting time between eruptions and the duration of the eruption for the Old Faithful geyser in Yellowstone National Park, Wyoming, USA). Look at the eruption times `r=faithful$eruptions`.

   (a) Draw a QQ-plot. Does the data look normally distributed.

   (b) Draw a QQ-plot for only those eruptions that last longer than 3.2 minutes.

   (c) Draw a QQ-plot for only those eruptions that less than 3.2 minutes.

   (d) What do you conclude about the data?

4. In R, let $n \geq 2$ denote an integer. Draw n samples from the $N(\mu, \sigma^2)$ distribution and calculate the statistic
$$\frac{\bar{X} - \mu}{S/\sqrt{n}}$$
Repeat this N times (with perhaps $n = 5$, $N = 10^4$). Compare the empirical distribution of the N sample means with the $N(0, 1)$ distribution and the t_{n-1} distribution. How large does n have to be before the distributions are hard to distinguish?

5. Consider drawing a sample of size n from Uniform[0,1] distribution and calculate the sample median.

   (a) Does the sample median converge in probability? If so, to what?

   (b) Does the sample median converge almost surely if you increase n by adding items to the sample one by one? [Hint: If a sequence of (non-necessarily independent) events $A_1, A_2, \ldots$ have probabilities $\mathbb{P}(A_n) \leq n^{-2}$ then only finitely many of the A_n will occur (almost surely).]

6. Derive the formula for the pdf of the k-th order statistic
$$f_{X_{(j)}}(x) = \frac{n!}{(j-1)!(n-j)!} f_X(x) \left[F_X(x)\right]^{j-1} \left[1 - F_X(x)\right]^{n-j}.$$
from the cdf
$$F_{X_{(j)}}(x) = \sum_{k=j}^{n} \binom{n}{k} \left[F_X(x)\right]^{k} \left[1 - F_X(x)\right]^{n-k}$$

7. Consider the Poisson(θ) distribution.

   (a) In R, sample 4 items with $\theta = 2.7$.

   (b) Plot the likelihood for θ in the range one to five.

   (c) Can you explain the position of the maximum in terms of the sample mean?

   (d) In R, sample 100 items with $\theta = 2.7$.

   (e) Plot the log of the likelihood for θ in the range one to five.

   (f) Can you explain the position of the maximum in terms of the sample mean?

Lab 4

Questions with a (*) next to them form part of homework 1. Due in 20/10/2014. [Q 3]

1. What is a statistic?

2. In R, load the iris dataset, and draw boxplots for each variable
```
data(iris)
boxplot(iris)
```
What do the different lines and circles represent?

3. [*] In R, load the faithful dataset (Waiting time between eruptions and the duration of the eruption for the Old Faithful geyser in Yellowstone National Park, Wyoming, USA). Look at the eruption times `r=faithful$eruptions`.

   (a) Draw a QQ-plot. Does the data look normally distributed.

   (b) Draw a QQ-plot for only those eruptions that last longer than 3.2 minutes.

   (c) Draw a QQ-plot for only those eruptions that less than 3.2 minutes.

   (d) What do you conclude about the data?

4. In R, let $n \geq 2$ denote an integer. Draw n samples from the $N(\mu, \sigma^2)$ distribution and calculate the statistic
$$\frac{\bar{X} - \mu}{S/\sqrt{n}}$$
Repeat this N times (with perhaps $n = 5$, $N = 10^4$). Compare the empirical distribution of the N sample means with the $N(0, 1)$ distribution and the t_{n-1} distribution. How large does n have to be before the distributions are hard to distinguish?

5. Consider drawing a sample of size n from Uniform[0,1] distribution and calculate the sample median.

   (a) Does the sample median converge in probability? If so, to what?

   (b) Does the sample median converge almost surely if you increase n by adding items to the sample one by one? [Hint: If a sequence of (non-necessarily independent) events $A_1, A_2, \ldots$ have probabilities $\mathbb{P}(A_n) \leq n^{-2}$ then only finitely many of the A_n will occur (almost surely).]

6. Derive the formula for the pdf of the k-th order statistic
$$f_{X_{(j)}}(x) = \frac{n!}{(j-1)!(n-j)!} f_X(x) \left[F_X(x)\right]^{j-1} \left[1 - F_X(x)\right]^{n-j}.$$
from the cdf
$$F_{X_{(j)}}(x) = \sum_{k=j}^{n} \binom{n}{k} \left[F_X(x)\right]^{k} \left[1 - F_X(x)\right]^{n-k}$$

7. Consider the Poisson(θ) distribution.

   (a) In R, sample 4 items with $\theta = 2.7$.

   (b) Plot the likelihood for θ in the range one to five.

   (c) Can you explain the position of the maximum in terms of the sample mean?

   (d) In R, sample 100 items with $\theta = 2.7$.

   (e) Plot the log of the likelihood for θ in the range one to five.

   (f) Can you explain the position of the maximum in terms of the sample mean?

Lab 4

Questions with a (*) next to them form part of homework 1. Due in 20/10/2014. [Q 3]

1. What is a statistic?

2. In R, load the iris dataset, and draw boxplots for each variable
   ```
   data(iris)
   boxplot(iris)
   ```
   What do the different lines and circles represent?

3. [*] In R, load the faithful dataset (Waiting time between eruptions and the duration of the eruption for the Old Faithful geyser in Yellowstone National Park, Wyoming, USA). Look at the eruption times `r=faithful$eruptions`.

   (a) Draw a QQ-plot. Does the data look normally distributed.

   (b) Draw a QQ-plot for only those eruptions that last longer than 3.2 minutes.

   (c) Draw a QQ-plot for only those eruptions that less than 3.2 minutes.

   (d) What do you conclude about the data?

4. In R, let $n \geq 2$ denote an integer. Draw n samples from the $N(\mu, \sigma^2)$ distribution and calculate the statistic
   $$\frac{\bar{X} - \mu}{S/\sqrt{n}}$$
   Repeat this N times (with perhaps $n = 5$, $N = 10^4$). Compare the empirical distribution of the N sample means with the $N(0, 1)$ distribution and the t_{n-1} distribution. How large does n have to be before the distributions are hard to distinguish?

5. Consider drawing a sample of size n from Uniform[0,1] distribution and calculate the sample median.

   (a) Does the sample median converge in probability? If so, to what?

   (b) Does the sample median converge almost surely if you increase n by adding items to the sample one by one? [Hint: If a sequence of (non-necessarily independent) events $A_1, A_2, \ldots$ have probabilities $\mathbb{P}(A_n) \leq n^{-2}$ then only finitely many of the A_n will occur (almost surely).]

6. Derive the formula for the pdf of the k-th order statistic
   $$f_{X_{(j)}}(x) = \frac{n!}{(j-1)!(n-j)!} f_X(x) \left[F_X(x)\right]^{j-1} \left[1 - F_X(x)\right]^{n-j}.$$
   from the cdf
   $$F_{X_{(j)}}(x) = \sum_{k=j}^{n} \binom{n}{k} \left[F_X(x)\right]^{k} \left[1 - F_X(x)\right]^{n-k}$$

7. Consider the Poisson(θ) distribution.

   (a) In R, sample 4 items with $\theta = 2.7$.

   (b) Plot the likelihood for θ in the range one to five.

   (c) Can you explain the position of the maximum in terms of the sample mean?

   (d) In R, sample 100 items with $\theta = 2.7$.

   (e) Plot the log of the likelihood for θ in the range one to five.

   (f) Can you explain the position of the maximum in terms of the sample mean?

Lab 4

Questions with a (*) next to them form part of homework 1. Due in 20/10/2014. [Q 3]

1. What is a statistic?

2. In R, load the iris dataset, and draw boxplots for each variable
```
data(iris)
boxplot(iris)
```
What do the different lines and circles represent?

3. [*] In R, load the faithful dataset (Waiting time between eruptions and the duration of the eruption for the Old Faithful geyser in Yellowstone National Park, Wyoming, USA). Look at the eruption times `r=faithful$eruptions`.

   (a) Draw a QQ-plot. Does the data look normally distributed.

   (b) Draw a QQ-plot for only those eruptions that last longer than 3.2 minutes.

   (c) Draw a QQ-plot for only those eruptions that less than 3.2 minutes.

   (d) What do you conclude about the data?

4. In R, let $n \geq 2$ denote an integer. Draw n samples from the $N(\mu, \sigma^2)$ distribution and calculate the statistic

$$\frac{\bar{X} - \mu}{S/\sqrt{n}}$$

Repeat this N times (with perhaps $n = 5$, $N = 10^4$). Compare the empirical distribution of the N sample means with the $N(0,1)$ distribution and the t_{n-1} distribution. How large does n have to be before the distributions are hard to distinguish?

5. Consider drawing a sample of size n from Uniform[0,1] distribution and calculate the sample median.

   (a) Does the sample median converge in probability? If so, to what?

   (b) Does the sample median converge almost surely if you increase n by adding items to the sample one by one? [Hint: If a sequence of (non-necessarily independent) events $A_1, A_2, \ldots$ have probabilities $\mathbb{P}(A_n) \leq n^{-2}$ then only finitely many of the A_n will occur (almost surely).]

6. Derive the formula for the pdf of the k-th order statistic

$$f_{X_{(j)}}(x) = \frac{n!}{(j-1)!(n-j)!} f_X(x) \left[F_X(x)\right]^{j-1} \left[1 - F_X(x)\right]^{n-j}.$$

from the cdf

$$F_{X_{(j)}}(x) = \sum_{k=j}^{n} \binom{n}{k} \left[F_X(x)\right]^k \left[1 - F_X(x)\right]^{n-k}$$

7. Consider the Poisson(θ) distribution.

   (a) In R, sample 4 items with $\theta = 2.7$.

   (b) Plot the likelihood for θ in the range one to five.

   (c) Can you explain the position of the maximum in terms of the sample mean?

   (d) In R, sample 100 items with $\theta = 2.7$.

   (e) Plot the log of the likelihood for θ in the range one to five.

   (f) Can you explain the position of the maximum in terms of the sample mean?

Lab 4

Questions with a (*) next to them form part of homework 1. Due in 20/10/2014. [Q 3]

1. What is a statistic?

2. In R, load the iris dataset, and draw boxplots for each variable
```
data(iris)
boxplot(iris)
```
What do the different lines and circles represent?

3. [*] In R, load the faithful dataset (Waiting time between eruptions and the duration of the eruption for the Old Faithful geyser in Yellowstone National Park, Wyoming, USA). Look at the eruption times `r=faithful$eruptions`.

   (a) Draw a QQ-plot. Does the data look normally distributed.

   (b) Draw a QQ-plot for only those eruptions that last longer than 3.2 minutes.

   (c) Draw a QQ-plot for only those eruptions that less than 3.2 minutes.

   (d) What do you conclude about the data?

4. In R, let $n \geq 2$ denote an integer. Draw n samples from the $N(\mu, \sigma^2)$ distribution and calculate the statistic
$$\frac{\bar{X} - \mu}{S/\sqrt{n}}$$
Repeat this N times (with perhaps $n = 5$, $N = 10^4$). Compare the empirical distribution of the N sample means with the $N(0, 1)$ distribution and the t_{n-1} distribution. How large does n have to be before the distributions are hard to distinguish?

5. Consider drawing a sample of size n from Uniform[0,1] distribution and calculate the sample median.

   (a) Does the sample median converge in probability? If so, to what?

   (b) Does the sample median converge almost surely if you increase n by adding items to the sample one by one? [Hint: If a sequence of (non-necessarily independent) events $A_1, A_2, \ldots$ have probabilities $\mathbb{P}(A_n) \leq n^{-2}$ then only finitely many of the A_n will occur (almost surely).]

6. Derive the formula for the pdf of the k-th order statistic
$$f_{X_{(j)}}(x) = \frac{n!}{(j-1)!(n-j)!} f_X(x) \left[F_X(x)\right]^{j-1} \left[1 - F_X(x)\right]^{n-j}.$$
from the cdf
$$F_{X_{(j)}}(x) = \sum_{k=j}^{n} \binom{n}{k} \left[F_X(x)\right]^k \left[1 - F_X(x)\right]^{n-k}$$

7. Consider the Poisson(θ) distribution.

   (a) In R, sample 4 items with $\theta = 2.7$.

   (b) Plot the likelihood for θ in the range one to five.

   (c) Can you explain the position of the maximum in terms of the sample mean?

   (d) In R, sample 100 items with $\theta = 2.7$.

   (e) Plot the log of the likelihood for θ in the range one to five.

   (f) Can you explain the position of the maximum in terms of the sample mean?

Lab 4

Questions with a (*) next to them form part of homework 1. Due in 20/10/2014. [Q 3]

1. What is a statistic?

2. In R, load the iris dataset, and draw boxplots for each variable
```
data(iris)
boxplot(iris)
```
What do the different lines and circles represent?

3. [*] In R, load the faithful dataset (Waiting time between eruptions and the duration of the eruption for the Old Faithful geyser in Yellowstone National Park, Wyoming, USA). Look at the eruption times `r=faithful$eruptions`.

   (a) Draw a QQ-plot. Does the data look normally distributed.

   (b) Draw a QQ-plot for only those eruptions that last longer than 3.2 minutes.

   (c) Draw a QQ-plot for only those eruptions that less than 3.2 minutes.

   (d) What do you conclude about the data?

4. In R, let $n \geq 2$ denote an integer. Draw n samples from the $N(\mu, \sigma^2)$ distribution and calculate the statistic
$$\frac{\bar{X} - \mu}{S/\sqrt{n}}$$
Repeat this N times (with perhaps $n = 5$, $N = 10^4$). Compare the empirical distribution of the N sample means with the $N(0, 1)$ distribution and the t_{n-1} distribution. How large does n have to be before the distributions are hard to distinguish?

5. Consider drawing a sample of size n from Uniform[0,1] distribution and calculate the sample median.

   (a) Does the sample median converge in probability? If so, to what?

   (b) Does the sample median converge almost surely if you increase n by adding items to the sample one by one? [Hint: If a sequence of (non-necessarily independent) events $A_1, A_2, \ldots$ have probabilities $\mathbb{P}(A_n) \leq n^{-2}$ then only finitely many of the A_n will occur (almost surely).]

6. Derive the formula for the pdf of the k-th order statistic
$$f_{X_{(j)}}(x) = \frac{n!}{(j-1)!(n-j)!} f_X(x) \left[F_X(x)\right]^{j-1} \left[1 - F_X(x)\right]^{n-j}.$$
from the cdf
$$F_{X_{(j)}}(x) = \sum_{k=j}^{n} \binom{n}{k} \left[F_X(x)\right]^k \left[1 - F_X(x)\right]^{n-k}$$

7. Consider the Poisson(θ) distribution.

   (a) In R, sample 4 items with $\theta = 2.7$.

   (b) Plot the likelihood for θ in the range one to five.

   (c) Can you explain the position of the maximum in terms of the sample mean?

   (d) In R, sample 100 items with $\theta = 2.7$.

   (e) Plot the log of the likelihood for θ in the range one to five.

   (f) Can you explain the position of the maximum in terms of the sample mean?

Lab 4

Questions with a (*) next to them form part of homework 1. Due in 20/10/2014. [Q 3]

1. What is a statistic?

2. In R, load the iris dataset, and draw boxplots for each variable
```
data(iris)
boxplot(iris)
```
   What do the different lines and circles represent?

3. [*] In R, load the faithful dataset (Waiting time between eruptions and the duration of the eruption for the Old Faithful geyser in Yellowstone National Park, Wyoming, USA). Look at the eruption times `r=faithful$eruptions`.

   (a) Draw a QQ-plot. Does the data look normally distributed.

   (b) Draw a QQ-plot for only those eruptions that last longer than 3.2 minutes.

   (c) Draw a QQ-plot for only those eruptions that less than 3.2 minutes.

   (d) What do you conclude about the data?

4. In R, let $n \geq 2$ denote an integer. Draw n samples from the $N(\mu, \sigma^2)$ distribution and calculate the statistic
$$\frac{\bar{X} - \mu}{S/\sqrt{n}}$$
   Repeat this N times (with perhaps $n = 5$, $N = 10^4$). Compare the empirical distribution of the N sample means with the $N(0, 1)$ distribution and the t_{n-1} distribution. How large does n have to be before the distributions are hard to distinguish?

5. Consider drawing a sample of size n from Uniform[0,1] distribution and calculate the sample median.

   (a) Does the sample median converge in probability? If so, to what?

   (b) Does the sample median converge almost surely if you increase n by adding items to the sample one by one? [Hint: If a sequence of (non-necessarily independent) events $A_1, A_2, \ldots$ have probabilities $\mathbb{P}(A_n) \leq n^{-2}$ then only finitely many of the A_n will occur (almost surely).]

6. Derive the formula for the pdf of the k-th order statistic
$$f_{X_{(j)}}(x) = \frac{n!}{(j-1)!(n-j)!} f_X(x) \left[F_X(x)\right]^{j-1} \left[1 - F_X(x)\right]^{n-j}.$$
   from the cdf
$$F_{X_{(j)}}(x) = \sum_{k=j}^{n} \binom{n}{k} \left[F_X(x)\right]^k \left[1 - F_X(x)\right]^{n-k}$$

7. Consider the Poisson(θ) distribution.

   (a) In R, sample 4 items with $\theta = 2.7$.

   (b) Plot the likelihood for θ in the range one to five.

   (c) Can you explain the position of the maximum in terms of the sample mean?

   (d) In R, sample 100 items with $\theta = 2.7$.

   (e) Plot the log of the likelihood for θ in the range one to five.

   (f) Can you explain the position of the maximum in terms of the sample mean?

Lab 4

Questions with a (*) next to them form part of homework 1. Due in 20/10/2014. [Q 3]

1. What is a statistic?

2. In R, load the iris dataset, and draw boxplots for each variable
```
data(iris)
boxplot(iris)
```
What do the different lines and circles represent?

3. [*] In R, load the faithful dataset (Waiting time between eruptions and the duration of the eruption for the Old Faithful geyser in Yellowstone National Park, Wyoming, USA). Look at the eruption times `r=faithful$eruptions`.

   (a) Draw a QQ-plot. Does the data look normally distributed.

   (b) Draw a QQ-plot for only those eruptions that last longer than 3.2 minutes.

   (c) Draw a QQ-plot for only those eruptions that less than 3.2 minutes.

   (d) What do you conclude about the data?

4. In R, let $n \geq 2$ denote an integer. Draw n samples from the $N(\mu, \sigma^2)$ distribution and calculate the statistic
$$\frac{\bar{X} - \mu}{S/\sqrt{n}}$$
Repeat this N times (with perhaps $n = 5$, $N = 10^4$). Compare the empirical distribution of the N sample means with the $N(0, 1)$ distribution and the t_{n-1} distribution. How large does n have to be before the distributions are hard to distinguish?

5. Consider drawing a sample of size n from Uniform[0,1] distribution and calculate the sample median.

   (a) Does the sample median converge in probability? If so, to what?

   (b) Does the sample median converge almost surely if you increase n by adding items to the sample one by one? [Hint: If a sequence of (non-necessarily independent) events $A_1, A_2, \ldots$ have probabilities $\mathbb{P}(A_n) \leq n^{-2}$ then only finitely many of the A_n will occur (almost surely).]

6. Derive the formula for the pdf of the k-th order statistic
$$f_{X_{(j)}}(x) = \frac{n!}{(j-1)!(n-j)!} f_X(x) \left[F_X(x)\right]^{j-1} \left[1 - F_X(x)\right]^{n-j}.$$
from the cdf
$$F_{X_{(j)}}(x) = \sum_{k=j}^{n} \binom{n}{k} \left[F_X(x)\right]^k \left[1 - F_X(x)\right]^{n-k}$$

7. Consider the Poisson(θ) distribution.

   (a) In R, sample 4 items with $\theta = 2.7$.

   (b) Plot the likelihood for θ in the range one to five.

   (c) Can you explain the position of the maximum in terms of the sample mean?

   (d) In R, sample 100 items with $\theta = 2.7$.

   (e) Plot the log of the likelihood for θ in the range one to five.

   (f) Can you explain the position of the maximum in terms of the sample mean?

Lab 4

Questions with a (*) next to them form part of homework 1. Due in 20/10/2014. [Q 3]

1. What is a statistic?

2. In R, load the iris dataset, and draw boxplots for each variable
```
data(iris)
boxplot(iris)
```
What do the different lines and circles represent?

3. [*] In R, load the faithful dataset (Waiting time between eruptions and the duration of the eruption for the Old Faithful geyser in Yellowstone National Park, Wyoming, USA). Look at the eruption times `r=faithful$eruptions`.

   (a) Draw a QQ-plot. Does the data look normally distributed.

   (b) Draw a QQ-plot for only those eruptions that last longer than 3.2 minutes.

   (c) Draw a QQ-plot for only those eruptions that less than 3.2 minutes.

   (d) What do you conclude about the data?

4. In R, let $n \geq 2$ denote an integer. Draw n samples from the $N(\mu, \sigma^2)$ distribution and calculate the statistic

$$\frac{\bar{X} - \mu}{S/\sqrt{n}}$$

Repeat this N times (with perhaps $n = 5$, $N = 10^4$). Compare the empirical distribution of the N sample means with the $N(0, 1)$ distribution and the t_{n-1} distribution. How large does n have to be before the distributions are hard to distinguish?

5. Consider drawing a sample of size n from Uniform[0,1] distribution and calculate the sample median.

   (a) Does the sample median converge in probability? If so, to what?

   (b) Does the sample median converge almost surely if you increase n by adding items to the sample one by one? [Hint: If a sequence of (non-necessarily independent) events $A_1, A_2, \ldots$ have probabilities $\mathbb{P}(A_n) \leq n^{-2}$ then only finitely many of the A_n will occur (almost surely).]

6. Derive the formula for the pdf of the k-th order statistic

$$f_{X_{(j)}}(x) = \frac{n!}{(j-1)!(n-j)!} f_X(x) \left[F_X(x)\right]^{j-1} \left[1 - F_X(x)\right]^{n-j}.$$

from the cdf

$$F_{X_{(j)}}(x) = \sum_{k=j}^{n} \binom{n}{k} \left[F_X(x)\right]^k \left[1 - F_X(x)\right]^{n-k}$$

7. Consider the Poisson(θ) distribution.

   (a) In R, sample 4 items with $\theta = 2.7$.

   (b) Plot the likelihood for θ in the range one to five.

   (c) Can you explain the position of the maximum in terms of the sample mean?

   (d) In R, sample 100 items with $\theta = 2.7$.

   (e) Plot the log of the likelihood for θ in the range one to five.

   (f) Can you explain the position of the maximum in terms of the sample mean?